IN THE SPECIFICATION

Please make the amendments to the Specification as indicated below:

Please replace the paragraph on page 4, lines 5-11 with the following paragraph:

C1

Herein disclosed is the MN gene, a cellular gene which is the endogenous component of the MaTu agent. A full-length cDNA sequence for the MN gene is shown in Figures 1A-1C [SEQ. ID. NO.: 1]. Figures 15A-15F provide a complete genomic sequence for MN [SEQ. ID. NO.: 5]. Figure 25 provides the sequence for a proposed MN promoter region [SEQ. ID. NO.: 27].

Please replace the paragraph on page 5, lines 18-26 with the following paragraph:

C2

Further, such isolated nucleic acids that encode MN proteins or polypeptides can also include the MN nucleic acids of the genomic clone shown in Figures 15A-15F, that is, SEQ. ID. NO.: 5, as well as sequences that hybridize to it or its complement under stringent conditions, or would hybridize to SEQ. ID. NO.: 5 or to its complement under such conditions, but for the degeneracy of the genetic code. Degenerate variants of SEQ. ID. NOS.: 1 and 5 are within the scope of the invention.

Please replace the paragraph on page 6, lines 11-12 with the following paragraph:

(a) a nucleic acid having the nucleotide sequence shown in Figures 15A-15F [SEQ. ID. NO.: 5] and its complement;

Please replace the paragraph on page 9, lines 10-23 with the following paragraph:

In HeLa and in tumorigenic HeLa x fibroblast hybrid (H/F-T) cells, MN protein is manifested as a "twin" protein p54/58N; it is glycosylated and forms disulfide-linked oligomers. As determined by electrophoresis upon reducing gels, MN proteins have molecular weights in the range of from about 40 kd to about 70 kd, preferably from about 45 kd to about 65 kd, more preferably from about 48 kd to about 58 kd. Upon non-reducing gels, MN proteins in the form of oligomers have molecular weights in the range of from about 145 kd to about 160 kd, preferably from about 150 to about 155 kd, still more preferably from about 152 to about 154 kd. A predicted amino acid sequence for a preferred MN protein of this invention is shown in Figures 1A-1C [SEQ. ID. NO. 2].

Please replace the paragraph at page 12, lines 6-23 with the following paragraph:

The invention further is directed to MN-specific antibodies, which can be used diagnostically/prognostically and may be used therapeutically. Preferred according to this invention are MN-specific antibodies reactive with the epitopes represented respectively by the amino acid sequences of the MN protein shown in Figures 1A-1C as follows: from AA 62 to AA 67 [SEQ. ID. NO.: 10]; from AA 55 to AA 60 [SEQ. ID. NO.: 11]; from AA 127 to AA 147 [SEQ. ID. NO.: 12]; from AA 36 to AA 51 [SEQ. ID. NO.: 13]; from AA 68 to AA 91 [SEQ. ID. NO.: 14]; from AA 279 to AA 291 [SEQ. ID. NO.: 15]; and from AA 435 to AA 450 [SEQ. ID. NO.: 16]. More preferred are antibodies reactive with epitopes represented by SEQ. ID. NOS.: 10, 11 and 12. Still more preferred are antibodies reactive with the epitopes represented by SEQ. ID NOS: 10 and 11, as for example, respectively Mabs M75 and MN12. Most preferred are monoclonal antibodies reactive with the epitope represented by SEQ. ID. NO.: 10.

Please replace the paragraph at page 15, lines 4-14 with the following paragraph:

This invention also concerns methods of treating neoplastic disease and/or pre-neoplastic disease comprising inhibiting the expression of MN genes by administering antisense

nucleic acid sequences that are substantially complementary to mRNA transcribed from MN genes. Said antisense nucleic acid sequences are those that hybridize to such mRNA under stringent hybridization conditions. Preferred are antisense nucleic acid sequences that are substantially complementary to sequences at the 5' end of the MN cDNA sequence shown in Figures 1A-1C. Preferably said antisense nucleic acid sequences are oligonucleotides.

Please replace the paragraph on page 23, lines 3-9 with the following paragraph:

There are twenty main amino acids, each of which is specified by a different arrangement of three adjacent nucleotides (triplet code or codon), and which are linked together in a specific order to form a characteristic protein. A three-letter or one-letter convention is used herein to identify said amino acids, as, for example, in Figures 1A-1C as follows:

Please replace the paragraph on page 24, lines 8-11 with the following paragraph:

Figures 1A-1C provide the nucleotide sequence for a full-length MN cDNA [SEQ. ID. NO.: 1] clone isolated as described herein. Figures 1A-1C also set forth the predicted amino acid sequence [SEQ. ID. NO.: 2] encoded by the cDNA.

Please replace the paragraph on page 26, lines 17-24 with the following paragraph:

Figures 11A and 11B (discussed in Example 8) graphically illustrate the results from radioimmunoprecipitation experiments with ^{125}I-GEX-3X-MN protein and different antibodies. The radioactive protein (15 x 10^3 cpm/tube) was precipitated with ascitic fluid or sera and SAC as follows: (A) ascites with MAb M75; (B) rabbit anti-MaTu serum; (C) normal rabbit serum; (D) human serum L8; (E) human serum KH; and (F) human serum M7.

C9

Please replace the paragraph on page 28, lines 5-8 with the following paragraph:

Figures 15A-15F provide a 10,898 bp complete genomic sequence of MN [SEQ. ID. NO.: 5]. The base count is as follows: 2654 A; 2739 C; 2645 G; and 2859 T. The 11 exons are shown in capital letters.

C10

Please replace the paragraph on page 30, lines 21-23 with the following paragraph:

Figure 23A-1 to 23C illustrate flow cytometric analyses of asynchronous cell populations of control and MN cDNA-transfected NIH 3T3 cells.

C11

Please replace the two paragraphs on page 36, lines 7-24 with the following paragraphs:

Examples herein show that MX and MN are two different entities, that can exist independently of each other. MX (LCMV) as an exogenous, transmissible agent can multiply in fibroblasts and in H/F-N hybrid cells which are not expressing MN-related proteins (Figures 6A and 6B). In such cells, MX does not induce the production of MN protein. MN protein can be produced in HeLa and other tumor cells even in the absence of MX as shown in Figures 6-9. However, MX is a potent inducer of MN-related protein in HeLa cells; it increases its production thirty times over the concentration observed in uninfected cells (Figures 7 and 12, Table 2 in Example 8, below).

MN Gene--Cloning and Sequencing

Figures 1A-1C provide the nucleotide sequence for a full-length MN cDNA clone isolated as described below [SEQ. ID. NO.: 1]. Figures 15A-15F provide a complete MN genomic sequence [SEQ. ID. NO.: 5]. Figure 25 shows the nucleotide sequence for a proposed MN promoter [SEQ. ID. NO.: 27].

Please replace the paragraph beginning on page 37, line 12 to page 38, line 9 with the following paragraph:

It is further understood that the nucleotide sequences herein described and shown in Figures 1A-1C, 15A-15F and 25,

represent only the precise structures of the cDNA, genomic and promoter nucleotide sequences isolated and described herein. It is expected that slightly modified nucleotide sequences will be found or can be modified by techniques known in the art to code for substantially similar or homologous MN proteins and polypeptides, for example, those having similar epitopes, and such nucleotide sequences and proteins/polypeptides are considered to be equivalents for the purpose of this invention. DNA or RNA having equivalent codons is considered within the scope of the invention, as are synthetic nucleic acid sequences that encode proteins/polypeptides homologous or substantially homologous to MN proteins/polypeptides, as well as those nucleic acid sequences that would hybridize to said exemplary sequences [SEQ. ID. NOS. 1, 5 and 27] under stringent conditions or that but for the degeneracy of the genetic code would hybridize to said cDNA nucleotide sequences under stringent hybridization conditions. Modifications and variations of nucleic acid sequences as indicated herein are considered to result in sequences that are substantially the same as the exemplary MN sequences and fragments thereof.

Please replace the paragraph on page 40, lines 2-10 with the following paragraph:

Attempts to isolate a full-length clone from the original cDNA library failed. Therefore, we performed a rapid amplification of cDNA ends (RACE) using MN-specific primers, R1 and R2, derived from the 5' region of the original cDNA clone. The RACE product was inserted into pBluescript, and the entire population of recombinant plasmids was sequenced with an MN-specific primer ODN1. In that way, we obtained a reliable sequence at the very 5' end of the MN cDNA as shown in **Figures 1A-1C** [SEQ. ID. NO.: 1].

Please replace Table 1 on page 45, lines 1-30 with the following Table 1:

TABLE 1

Exon-Intron Structure of the Human MN Gene

Exon	Size	Genomic Position**	SEQ ID NO	5'splice donor	SEQ ID No
1	445	*3507-3951	28	AGAAG gtaagt	67
2	30	5126-5155	29	TGGAG gtgaga	68
3	171	5349-5519	30	CAGTC gtgagg	69
4	143	5651-5793	31	CCGAG gtgagc	70
5	93	5883-5975	32	TGGAG gtacca	71
6	67	7376-7442	33	GGAAG gtcagt	72
7	158	8777-8934	34	AGCAG gtgggc	73
8	145	9447-9591	35	GCCAG gtacag	74
9	27	9706-9732	36	TGCTG gtgagt	75
10	82	10350-10431	37	CACAG gtatta	76
11	191	10562-10752	38	ATAAT end	

Intron	Size	Genomic Position**	SEQ ID NO	3'splice acceptor	SEQ ID NO
1	1174	3952-5125	39	atacag GGGAT	77
2	193	5156-5348	40	ccccag GCGAC	78
3	131	5520-5650	41	acgcag TGCAA	79
4	89	5794-5882	42	tttcag ATCCA	80
5	1400	5976-7375	43	ccccag GAGGG	81
6	1334	7443-8776	44	tcacag GCTCA	82
7	512	8935-9446	45	ccctag CTCCA	83
8	114	9592-9705	46	ctccag TCCAG	84
9	617	9733-10349	47	tcgcag GTGACA	85
10	130	10432-10561	48	acacag AAGGG	86

** positions are related to nt numbering in whole genomic sequence including the 5' flanking region [Figures 15A-15F]

* number corresponds to transcription initiation site determined below by RNase protection assay

Please replace the two paragraphs beginning on page 55, line 2 to page 56, line 6 with the following paragraphs:

The ORF of the MN cDNA shown in Figures 1A-1C have the coding capacity for a 459 amino acid protein with a calculated molecular weight of 49.7 kd. MN protein has an estimated pI of about 4. As assessed by amino acid sequence analysis, the deduced primary structure of the MN protein can be divided into four distinct regions. The initial hydrophobic region of 37 amino acids (AA) corresponds to a signal peptide. The mature protein has an N-terminal part of 377 AA, a hydrophobic transmembrane segment of 20 AA and a C-terminal region of 25 AA. Alternatively, the MN protein can be viewed as having five domains as follows: (1) a signal peptide [amino acids (AA) 1-37; SEQ. ID. NO.: 6]; (2) a region of homology to collagen alpha1 chain (AA 38-135; SEQ. ID. NO.: 50); (3) a carbonic anhydrase domain (AA 136-391; SEQ. ID. NO.: 51); (4) a transmembrane region (AA 414-433; SEQ. ID. NO.: 52); and (5) an intracellular C terminus (AA 435-459; SEQ. ID. NO.: 53). (The AA numbers are keyed to Figures 1A-1C.)

More detailed insight into MN protein primary structure disclosed the presence of several consensus sequences. One potential N-glycosylation site was found at position 346 of Figures 1A-1C. That feature, together with a predicted membrane-

spanning region are consistent with the results, in which MN was shown to be an N-glycosylated protein localized in the plasma membrane. MN protein sequence deduced from cDNA was also found to contain seven S/TPXX sequence elements [SEQ. ID. NOS.: 25 AND 26] (one of them is in the signal peptide) defined by Suzuki, J. Mol. Biol., 207: 61-84 (1989) as motifs frequently found in gene regulatory proteins. However, only two of them are composed of the suggested consensus amino acids.

Please replace the two paragraphs beginning on page 57, line 6 to page 58, line 2 with the following paragraphs:

The MN gene was found to clearly be a novel sequence derived from the human genome. Searches for amino acid sequence similarities in protein databases revealed as the closest homology a level of sequence identity (38.9% in 256 AA or 44% in an 170 AA overlap) between the central part of the MN protein [AAs 136-391 (SEQ. ID. NO: 51)] or 221-390 [SEQ. ID. NO.: 54] of Figures 1A-1C and carbonic anhydrases (CA). However, the overall sequence homology between the cDNA MN sequence and cDNA sequences encoding different CA isoenzymes is in a homology range of 48-50% which is considered by ones in the art to be low. Therefore, the MN cDNA sequence is not closely related to any CA cDNA sequences.

Only very closely related nt sequences having a homology of at least 80-90% would hybridize to each other under stringent conditions. A sequence comparison of the MN cDNA sequence shown in Figures 1A-1C and a corresponding cDNA of the human carbonic anhydrase II (CA II) showed that there are no stretches of identity between the two sequences that would be long enough to allow for a segment of the CA II cDNA sequence having 50 or more nucleotides to hybridize under stringent hybridization conditions to the MN cDNA or vice versa.

Please replace the two paragraphs beginning on page 59, line 18 to page 60, line 4 with the following paragraphs:

The phrase "MN proteins and/or polypeptides" (MN proteins/polypeptides) is herein defined to mean proteins and/or polypeptides encoded by an MN gene or fragments thereof. An exemplary and preferred MN protein according to this invention has the deduced amino acid sequence shown in Figures 1A-1C. Preferred MN proteins/polypeptides are those proteins and/or polypeptides that have substantial homology with the MN protein shown in Figures 1A-1C. For example, such substantially homologous MN proteins/polypeptides are those that are reactive with the MN-specific antibodies of this invention, preferably the Mabs M75, MN12, MN9 and MN7 or their equivalents.

Please replace the paragraph on page 62, lines 4-13 with the following paragraph:

A representative method to prepare the MN proteins shown in Figures 1A-1C or fragments thereof would be to insert the full-length or an appropriate fragment of MN cDNA into an appropriate expression vector as exemplified below. The fusion protein GEX-3X-MN expressed from XL1-Blue cells is nonglycosylated. Representative of a glycosylated, recombinantly produced MN protein is the MN 20-19 protein expressed from insect cells. The MN 20-19 protein was also expressed in a nonglycosylated form in E. coli using the expression plasmid pEt-22b [Novagen].

Please replace the paragraph beginning on page 69, line 13 to page 70, line 3 with the following paragraph:

Another representative, recombinantly produced MN protein of this invention is the MN 20-19 protein which, when produced in baculovirus-infected Sf9 cells [Spodoptera frugiperda cells; Clontech; Palo Alto, CA (USA)], is glycosylated. The MN 20-19 protein misses the putative signal peptide (AAs 1-37) of SEQ. ID. NO.: 6 (Figures 1A-1C), has a methionine (Met) at the N-terminus for expression, and a Leu-Glu-His-His-His-His-His-His [SEQ. ID NO.: 22] added to the C-terminus for purification. In order to insert the portion of the MN coding sequence for the

GEX-3X-MN fusion protein into alternate expression systems, a set of primers for PCR was designed. The primers were constructed to provide restriction sites at each end of the coding sequence, as well as in-frame start and stop codons. The sequences of the primers, indicating restriction enzyme cleavage sites and expression landmarks, are shown below.

Please replace the paragraph on page 81, lines 11-25 with the following paragraph:

Nucleic acid probes of this invention are those comprising sequences that are complementary or substantially complementary to the MN cDNA sequence shown in Figures 1A-1C or to other MN gene sequences, such as, the complete genomic sequence of Figures 15A-15F [SEQ. ID. NO.: 5] and the putative promoter sequence [SEQ. ID. NO.: 27 of Figure 25]. The phrase "substantially complementary" is defined herein to have the meaning as it is well understood in the art and, thus, used in the context of standard hybridization conditions. The stringency of hybridization conditions can be adjusted to control the precision of complementarity. Exemplary are the stringent hybridization conditions used in Examples 11 and 12. Two nucleic acids are, for example, substantially complementary to each

other, if they hybridize to each other under such stringent hybridization conditions.

Please replace the paragraph on page 83, lines 1-10 with the following paragraph:

However, nucleic acid probes of this invention need not hybridize to a coding region of MN. For example, nucleic acid probes of this invention may hybridize partially or wholly to a non-coding region of the genomic sequence shown in Figures 15A-15F [SEQ. ID. NO.: 5]. Conventional technology can be used to determine whether fragments of SEQ. ID. NO.: 5 or related nucleic acids are useful to identify MN nucleic acid sequences. [See, for example, Benton and Davis, supra and Fuscoe et al., supra.]

Please replace the table on page 84, lines 1-12 with the following table:

Region of Homology within MN Genomic Sequence [SEQ. ID. NO.: 5; Figures 15A-15F]	SEQ. ID. NOS.	% Homology to Entire Alu-J Sequence
921-1212	59	89.1%
2370-2631	60	78.6%
4587-4880	61	90.1%
6463-6738	62	85.4%
7651-7939	63	91.0%
9020-9317	64	69.8%
		% Homology to One Half of Alu-J Sequence
8301-8405	65	88.8%
10040-10122	66	73.2%.

Please replace the paragraph on page 98, lines 9-18 with the following paragraph:

Anti-peptide antibodies are also made by conventional methods in the art as described in European Patent Publication No. 44,710 (published Jan. 27, 1982). Briefly, such anti-peptide antibodies are prepared by selecting a peptide from an MN amino acid sequence as from Figures 1A-1C, chemically synthesizing it, conjugating it to an appropriate immunogenic protein and injecting it into an appropriate animal, usually a rabbit or a

mouse; then, either polyclonal or monoclonal antibodies are made, the latter by a Kohler-Milstein procedure, for example.

Please replace the paragraph on page 102, lines 14-19 with the following paragraph:

Mab M75 recognizes both the nonglycosylated GEX-3X-MN fusion protein and native MN protein as expressed in CGL3 cells equally well. Mab M75 was shown by epitope mapping to be reactive with the epitope represented by the amino acid sequence from AA 62 to AA 67 [SEQ. ID. NO.: 10] of the MN protein shown in Figures 1A-1C.

Please replace the paragraph on page 104, lines 1-5 with the following paragraph:

Mab MN9. Monoclonal antibody MN9 (Mab MN9) reacts to the same epitope as Mab M75, represented by the sequence from AA 62 to AA 67 [SEQ. ID. NO.: 10] of the Figures 1A-1C MN protein. As Mab M75, Mab MN9 recognizes both the GEX-3X-MN fusion protein and native MN protein equally well.

Please replace the two paragraphs beginning on page 104, line 14 to page 105, line 10 with the following paragraphs:

Mab MN12. Monoclonal antibody MN12 (Mab MN12) is produced by the mouse lymphocytic hybridoma MN 12.2.2 which was

deposited under ATCC Designation HB 11647 on June 9, 1994 at the American Type Culture Collection (ATCC) at 10801 University Blvd., Manassas, Virginia 20110-2209 (USA). Antibodies corresponding to Mab MN12 can also be made, analogously to the method outlined above for Mab MN9, by screening a series of antibodies prepared against an MN protein/polypeptide, against the peptide representing the epitope for Mab MN12. That peptide is AA 55 - AA 60 of Figures 1A-1C [SEQ. ID. NO.: 11]. The Novatope system could also be used to find antibodies specific for said epitope.

Mab MN7. Monoclonal antibody MN7 (Mab MN7) was selected from mabs prepared against nonglycosylated GEX-3X-MN as described above. It recognizes the epitope on MN represented by the amino acid sequence from AA 127 to AA 147 [SEQ. ID. NO.: 12] of the Figures 1A-1C MN protein. Analogously to methods described above for Mabs MN9 and MN12, mabs corresponding to Mab MN7 can be prepared by selecting mabs prepared against an MN protein/polypeptide that are reactive with the peptide having SEQ. ID. NO.: 12, or by the stated alternative means.

Please replace the paragraph on page 109, lines 1-11 with the following paragraph:

Preferred antisense oligonucleotides according to this invention are gene-specific ODNs or oligonucleotides complementary to the 5' end of MN mRNA. Particularly preferred are the 29-mer ODN1 and 19-mer ODN2 for which the sequences are provided in Example 10, infra. Those antisense ODNs are representative of the many antisense nucleic acid sequences that can function to inhibit MN gene expression. Ones of ordinary skill in the art could determine appropriate antisense nucleic acid sequences, preferably antisense oligonucleotides, from the nucleic acid sequences of Figures 1A-1C and 15A-15F.

Please replace the paragraph beginning on page 121, line 14 to page 122, line 2 with the following paragraph:

As shown in Figures 6A and 6B discussed below in Example 5, MX antigen was found to be present in MaTu-infected fibroblasts. In Zavada and Zavadova, supra, it was reported that a p58 band from MX-infected fibroblasts could not be detected by RIP with rabbit anti-MaTu serum. That serum contains more antibodies to MX than to MN antigen. The discrepancy can be explained by the extremely slow spread of MX in infected cultures. The results reported in Zavada and Zavadova, supra were from fibroblasts tested 6 weeks after infection, whereas the later testing was 4 months after infection. We have found by

immunoblots that MX can be first detected in both H/F-N and H/F-T hybrids after 4 weeks, in HeLa cells after six weeks and in fibroblasts only 10 weeks after infection.

Please replace the three paragraphs beginning on page 122, line 5 to page 123, line 9 with the following paragraphs:

Figures 6A and 6B graphically illustrate the expression of MN- and MX- specific proteins in human fibroblasts, in HeLa cells and in H/F-N and H/F-T hybrid cells, and contrasts the expression in MX-infected and uninfected cells. Cells were infected with MX by co-cultivation with mitomycin C-treated MX-infected HeLa. The infected and uninfected cells were grown for three passages in dense cultures. About four months after infection, the infected cells concurrently with uninfected cells were grown in petri dishes to produce dense monolayers.

A radioimmunoassay was performed directly in confluent petri dish (5 cm) culture of cells, fixed with methanol essentially as described in Example 3, supra. The monolayers were fixed with methanol and treated with ^{125}I-labeled MAbs M67 (specific for exogenous MX antigen) or M75 (specific for endogenous MN antigen) at 6 x 10^4 cpm/dish. The bound radioactivity was measured; the results are shown in Figures 6A and 6B.

Figures 6A and 6B show that MX was transmitted to all four cell lines tested, that is, to human embryo fibroblasts, to HeLa and to both H/F-N and H/F-T hybrids; at the same time, all four uninfected counterpart cell lines were MX-negative (top graph of Figures 6A and 6B). MN antigens are shown to be present in both MX-infected and uninfected HeLa and H/F-T cells, but not in the fibroblasts (bottom graph of Figures 6A and 6B). No MN antigen was found in the control H/F-N, and only a minimum increase over background of MN antigen was found in MaTu infected H/F-N. Thus, it was found that in the hybrids, expression of MN antigen very strongly correlates with tumorigenicity.

Please replace the paragraph on page 129, lines 11-23 with the following paragraph:

Titration of antibodies to MN antigen is shown in Figures 11A and 11B. Ascitic fluid from a mouse carrying M75 hybridoma cells (A) is shown to have a 50% end-point at dilution $1:1.4 \times 10^{-6}$. At the same time, ascitic fluids with MAbs specific for MX protein (M16 and M67) showed no precipitation of ^{125}I-labeled GEX-3X-MN even at dilution 1:200 (result not shown). Normal rabbit serum (C) did not significantly precipitate the MN antigen; rabbit anti-MaTu serum (B), obtained after immunization with live MX-infected HeLa cells, precipitated 7% of radioactive

MN protein, when diluted 1:200. The rabbit anti-MaTu serum is shown by immunoblot in Example 4 (above) to precipitate both MX and MN proteins.

Please replace the paragraph beginning on page 132, line 15 to page 133, line 11 with the following paragraph:

Ultrathin sections of control and of MX-infected HeLa cells are shown in Figures 13A-13D. Those immuno-electron micrographs demonstrate the location of MN antigen in the cells, and in addition, the striking ultrastructural differences between control and MX-infected HeLa. A control HeLa cell (Figure 13A) is shown to have on its surface very little MN antigen, as visualised with gold beads. The cell surface is rather smooth, with only two little protrusions. No mitochondria can be seen in the cytoplasm. In contrast, MX-infected HeLa cells (Figures 13B and 13C show the formation of abundant, dense filamentous protrusions from their surfaces. Most of the MN antigen is located on those filaments, which are decorated with immunogold. The cytoplasm of MX-infected HeLa contains numerous mitochondria (Figure 13C). Figure 13D demonstrates the location of MN antigen in the nucleus: some of the MN antigen is in nucleoplasm (possibly linked to chromatin), but a higher concentration of the MN antigen is in the nucleoli. Again, the surface of normal HeLa

(panels A and E of Figure 13) is rather smooth whereas MX-infected HeLa cells have on their surface, numerous filaments and "blebs". Some of the filaments appear to form bridges connecting them to adjacent cells.

Please replace the paragraph on page 151, lines 11-23 with the following paragraph:

The MN-expressing NIH 3T3 cells displayed spindle-shaped morphology, and increased refractility; they were less adherent to the solid support and smaller in size. The control (mock transfected cells) had a flat morphology, similar to parental NIH 3T3 cells. In contrast to the control cells that were aligned and formed a monolayer with an ordered pattern, the cells expressing MN lost the capacity for growth arrest and grew chaotically on top of one another (Figures 22A-22D. Correspondingly, the MN-expressing cells were able to reach significantly higher (more than 2x) saturation densities (Table 4) and were less dependent on growth factors than the control cells (Figures 22G and 22H).

Please replace the three paragraphs beginning on page 153, line 11 to page 154, line 13 with the following paragraphs:

Flow cytometric analyses of asynchronous cell populations. For the results shown in Figures 23A-1 and 23A-2,

cells that had been grown in dense culture were plated at 1 x 10^6 cells per 60 mm dish. Four days later, the cells were collected by trypsinization, washed, resuspended in PBS, fixed by dropwise addition of 70% ethanol and stained by propidium iodine solution containing RNase. Analysis was performed by FACStar using DNA cell cycle analysis software [Becton Dickinson; Franklin Lakes, NJ (USA)].

For the analyses shown in Figures 23B-1, 23B-2 and 23C, exponentially growing cells were plated at 5 x 10^5 cells per 60 mm dish and analysed as above 2 days later. Forward light scatter was used for the analysis of relative cell sizes. The data were evaluated using Kolmogorov-Smirnov test [Young, J. Histochem. Cytochem., 25: 935 (1977)]. D is the maximum difference between summation curves derived from histograms. D/s(n) is a value which indicates the similarity of the compared curves (it is close to zero when curves are similar).

The flow cytometric analyses revealed that clonal populations constitutively expressing MN protein showed a decreased percentage of cells in G1 phase and an increased percentage of cells in G2-M phases. Those differences were more striking in cell populations grown throughout three passages in high density cultures Figures 23A-1 and 23A-2], than in exponentially growing subconfluent cells Figures 23B-1 and 23B-2.

That observation supports the idea that MN protein has the capacity to perturb contact inhibition.

On page 159, after line 12, please insert the following Sequence Listing.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Zavada, Jan
Pastorekova, Silvia
Pastorek, Jaromir

(ii) TITLE OF INVENTION: MN Gene and Protein

(iii) NUMBER OF SEQUENCES: 86

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Leona L. Lauder
(B) STREET: 369 Pine Street
(C) CITY: San Francisco
(D) STATE: California
(E) COUNTRY: USA
(F) ZIP: 94104

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 09/772,719
(B) FILING DATE: 01-30-2001
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/485,049
(B) FILING DATE: 07-JUN-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Lauder, Leona L.
(B) REGISTRATION NUMBER: 30,863
(C) REFERENCE/DOCKET NUMBER: D-0021.3A-2

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 415-981-2034
(B) TELEFAX: 415-981-0332

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1522 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ACAGTCAGCC GCATGGCTCC CCTGTGCCCC AGCCCCTGGC TCCCTCTGTT GATCCCGGCC 60

CCTGCTCCAG GCCTCACTGT GCAACTGCTG CTGTCACTGC TGCTTCTGAT GCCTGTCCAT 120

CCCCAGAGGT TGCCCCGGAT GCAGGAGGAT TCCCCCTTGG GAGGAGGCTC TTCTGGGGAA 180

GATGACCCAC TGGGCGAGGA GGATCTGCCC AGTGAAGAGG ATTCACCCAG AGAGGAGGAT 240

CCACCCGGAG AGGAGGATCT ACCTGGAGAG GAGGATCTAC CTGGAGAGGA GGATCTACCT 300

GAAGTTAAGC CTAAATCAGA AGAAGAGGGC TCCCTGAAGT TAGAGGATCT ACCTACTGTT 360

GAGGCTCCTG GAGATCCTCA AGAACCCCAG AATAATGCCC ACAGGGACAA AGAAGGGGAT 420

GACCAGAGTC ATTGGCGCTA TGGAGGCGAC CCGCCCTGGC CCCGGGTGTC CCCAGCCTGC 480

GCGGGCCGCT TCCAGTCCCC GGTGGATATC CGCCCCCAGC TCGCCGCCTT CTGCCCGGCC 540

CTGCGCCCCC TGGAACTCCT GGGCTTCCAG CTCCCGCCGC TCCCAGAACT GCGCCTGCGC 600

AACAATGGCC ACAGTGTGCA ACTGACCCTG CCTCCTGGGC TAGAGATGGC TCTGGGTCCC 660

GGGCGGGAGT ACCGGGCTCT GCAGCTGCAT CTGCACTGGG GGGCTGCAGG TCGTCCGGGC 720

TCGGAGCACA CTGTGGAAGG CCACCGTTTC CCTGCCGAGA TCCACGTGGT TCACCTCAGC 780

ACCGCCTTTG CCAGAGTTGA CGAGGCCTTG GGGCGCCCGG GAGGCCTGGC CGTGTTGGCC 840

GCCTTTCTGG AGGAGGGCCC GGAAGAAAAC AGTGCCTATG AGCAGTTGCT GTCTCGCTTG 900

GAAGAAATCG CTGAGGAAGG CTCAGAGACT CAGGTCCCAG GACTGGACAT ATCTGCACTC 960

CTGCCCTCTG ACTTCAGCCG CTACTTCCAA TATGAGGGGT CTCTGACTAC ACCGCCCTGT 1020

GCCCAGGGTG TCATCTGGAC TGTGTTTAAC CAGACAGTGA TGCTGAGTGC TAAGCAGCTC 1080

CACACCCTCT CTGACACCCT GTGGGGACCT GGTGACTCTC GGCTACAGCT GAACTTCCGA 1140

GCGACGCAGC CTTTGAATGG GCGAGTGATT GAGGCCTCCT TCCCTGCTGG AGTGGACAGC 1200

AGTCCTCGGG CTGCTGAGCC AGTCCAGCTG AATTCCTGCC TGGCTGCTGG TGACATCCTA 1260

GCCCTGGTTT TTGGCCTCCT TTTTGCTGTC ACCAGCGTCG CGTTCCTTGT GCAGATGAGA 1320

AGGCAGCACA GAAGGGGAAC CAAAGGGGGT GTGAGCTACC GCCCAGCAGA GGTAGCCGAG 1380

ACTGGAGCCT AGAGGCTGGA TCTTGGAGAA TGTGAGAAGC CAGCCAGAGG CATCTGAGGG 1440

GGAGCCGGTA ACTGTCCTGT CCTGCTCATT ATGCCACTTC CTTTTAACTG CCAAGAAATT 1500

TTTTAAAATA AATATTTATA AT 1522

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 459 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(A) DESCRIPTION: First 37 amino acids represent

signal peptide, and remaining amino acids represent mature protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met Ala Pro Leu Cys Pro Ser Pro Trp Leu Pro Leu Leu Ile Pro Ala
        -35                 -30                 -25

Pro Ala Pro Gly Leu Thr Val Gln Leu Leu Leu Ser Leu Leu Leu Leu
    -20                 -15                 -10


Met Pro Val His Pro Gln Arg Leu Pro Arg Met Gln Glu Asp Ser Pro
-5                  1                   5                  10

Leu Gly Gly Gly Ser Ser Gly Glu Asp Asp Pro Leu Gly Glu Glu Asp
                15                  20                  25

Leu Pro Ser Glu Glu Asp Ser Pro Arg Glu Glu Asp Pro Pro Gly Glu
        30                  35                  40

Glu Asp Leu Pro Gly Glu Glu Asp Leu Pro Gly Glu Glu Asp Leu Pro
    45                  50                  55

Glu Val Lys Pro Lys Ser Glu Glu Glu Gly Ser Leu Lys Leu Glu Asp
60                  65                  70                  75

Leu Pro Thr Val Glu Ala Pro Gly Asp Pro Gln Glu Pro Gln Asn Asn
                80                  85                  90

Ala His Arg Asp Lys Glu Gly Asp Asp Gln Ser His Trp Arg Tyr Gly
            95                  100                 105

Gly Asp Pro Pro Trp Pro Arg Val Ser Pro Ala Cys Ala Gly Arg Phe
        110                 115                 120

Gln Ser Pro Val Asp Ile Arg Pro Gln Leu Ala Ala Phe Cys Pro Ala
    125                 130                 135

Leu Arg Pro Leu Glu Leu Leu Gly Phe Gln Leu Pro Pro Leu Pro Glu
140                 145                 150                 155

Leu Arg Leu Arg Asn Asn Gly His Ser Val Gln Leu Thr Leu Pro Pro
                160                 165                 170
```

```
Gly Leu Glu Met Ala Leu Gly Pro Gly Arg Glu Tyr Arg Ala Leu Gln
                175                 180                 185

Leu His Leu His Trp Gly Ala Ala Gly Arg Pro Gly Ser Glu His Thr
            190                 195                 200

Val Glu Gly His Arg Phe Pro Ala Glu Ile His Val Val His Leu Ser
    205                 210                 215

Thr Ala Phe Ala Arg Val Asp Glu Ala Leu Gly Arg Pro Gly Gly Leu
220                 225                 230                 235

Ala Val Leu Ala Ala Phe Leu Glu Glu Gly Pro Glu Glu Asn Ser Ala
                240                 245                 250

Tyr Glu Gln Leu Leu Ser Arg Leu Glu Glu Ile Ala Glu Glu Gly Ser
                255                 260                 265

Glu Thr Gln Val Pro Gly Leu Asp Ile Ser Ala Leu Leu Pro Ser Asp
            270                 275                 280

Phe Ser Arg Tyr Phe Gln Tyr Glu Gly Ser Leu Thr Thr Pro Pro Cys
    285                 290                 295

Ala Gln Gly Val Ile Trp Thr Val Phe Asn Gln Thr Val Met Leu Ser
300                 305                 310                 315

Ala Lys Gln Leu His Thr Leu Ser Asp Thr Leu Trp Gly Pro Gly Asp
                320                 325                 330

Ser Arg Leu Gln Leu Asn Phe Arg Ala Thr Gln Pro Leu Asn Gly Arg
                335                 340                 345

Val Ile Glu Ala Ser Phe Pro Ala Gly Val Asp Ser Ser Pro Arg Ala
            350                 355                 360

Ala Glu Pro Val Gln Leu Asn Ser Cys Leu Ala Ala Gly Asp Ile Leu
    365                 370                 375

Ala Leu Val Phe Gly Leu Leu Phe Ala Val Thr Ser Val Ala Phe Leu
380                 385                 390                 395

Val Gln Met Arg Arg Gln His Arg Arg Gly Thr Lys Gly Gly Val Ser
                400                 405                 410
```

Tyr Arg Pro Ala Glu Val Ala Glu Thr Gly Ala
415 420

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CGCCCAGTGG GTCATCTTCC CCAGAAGAG 29

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGAATCCTCC TGCATCCGG 19

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10898 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGATCCTGTT GACTCGTGAC CTTACCCCCA ACCCTGTGCT CTCTGAAACA TGAGCTGTGT 60

CCACTCAGGG TTAAATGGAT TAAGGGCGGT GCAAGATGTG CTTTGTTAAA CAGATGCTTG 120

AAGGCAGCAT GCTCGTTAAG AGTCATCACC AATCCCTAAT CTCAAGTAAT CAGGGACACA 180

AACACTGCGG AAGGCCGCAG GGTCCTCTGC CTAGGAAAAC CAGAGACCTT TGTTCACTTG 240

TTTATCTGAC CTTCCCTCCA CTATTGTCCA TGACCCTGCC AAATCCCCCT CTGTGAGAAA 300

CACCCAAGAA TTATCAATAA AAAAATAAAT TTAAAAAAAA AATACAAAAA AAAAAAAAAA 360

AAAAAAAAAA GACTTACGAA TAGTTATTGA TAAATGAATA GCTATTGGTA AAGCCAAGTA 420

AATGATCATA TTCAAAACCA GACGGCCATC ATCACAGCTC AAGTCTACCT GATTTGATCT 480

CTTTATCATT GTCATTCTTT GGATTCACTA GATTAGTCAT CATCCTCAAA ATTCTCCCCC 540

AAGTTCTAAT TACGTTCCAA ACATTTAGGG GTTACATGAA GCTTGAACCT ACTACCTTCT 600

TTGCTTTTGA GCCATGAGTT GTAGGAATGA TGAGTTTACA CCTTACATGC TGGGGATTAA 660

TTTAAACTTT ACCTCTAAGT CAGTTGGGTA GCCTTTGGCT TATTTTTGTA GCTAATTTTG 720

TAGTTAATGG ATGCACTGTG AATCTTGCTA TGATAGTTTT CCTCCACACT TTGCCACTAG 780

GGGTAGGTAG GTACTCAGTT TTCAGTAATT GCTTACCTAA GACCCTAAGC CCTATTTCTC 840

TTGTACTGGC CTTTATCTGT AATATGGGCA TATTTAATAC AATATAATTT TTGGAGTTTT 900

TTTGTTTGTT TGTTTGTTTG TTTTTTTGAG ACGGAGTCTT GCATCTGTCA TGCCCAGGCT 960

GGAGTAGCAG TGGTGCCATC TCGGCTCACT GCAAGCTCCA CCTCCCGAGT TCACGCCATT 1020

TTCCTGCCTC AGCCTCCCGA GTAGCTGGGA CTACAGGCGC CCGCCACCAT GCCCGGCTAA 1080

TTTTTTGTAT TTTTGGTAGA GACGGGGTTT CACCGTGTTA GCCAGAATGG TCTCGATCTC 1140

CTGACTTCGT GATCCACCCG CCTCGGCCTC CCAAAGTTCT GGGATTACAG GTGTGAGCCA 1200

CCGCACCTGG CCAATTTTTT GAGTCTTTTA AAGTAAAAAT ATGTCTTGTA AGCTGGTAAC 1260

TATGGTACAT TTCCTTTTAT TAATGTGGTG CTGACGGTCA TATAGGTTCT TTTGAGTTTG 1320

GCATGCATAT GCTACTTTTT GCAGTCCTTT CATTACATTT TTCTCTCTTC ATTTGAAGAG 1380

CATGTTATAT CTTTTAGCTT CACTTGGCTT AAAAGGTTCT CTCATTAGCC TAACACAGTG 1440

TCATTGTTGG TACCACTTGG ATCATAAGTG GAAAAACAGT CAAGAAATTG CACAGTAATA 1500

CTTGTTTGTA AGAGGGATGA TTCAGGTGAA TCTGACACTA AGAAACTCCC CTACCTGAGG 1560

TCTGAGATTC CTCTGACATT GCTGTATATA GGCTTTTCCT TTGACAGCCT GTGACTGCGG 1620

ACTATTTTTC TTAAGCAAGA TATGCTAAAG TTTTGTGAGC CTTTTTCCAG AGAGAGGTCT 1680

CATATCTGCA TCAAGTGAGA ACATATAATG TCTGCATGTT TCCATATTTC AGGAATGTTT 1740

GCTTGTGTTT TATGCTTTTA TATAGACAGG GAAACTTGTT CCTCAGTGAC CCAAAAGAGG 1800

TGGGAATTGT TATTGGATAT CATCATTGGC CCACGCTTTC TGACCTTGGA AACAATTAAG 1860

GGTTCATAAT CTCAATTCTG TCAGAATTGG TACAAGAAAT AGCTGCTATG TTTCTTGACA 1920

TTCCACTTGG TAGGAAATAA GAATGTGAAA CTCTTCAGTT GGTGTGTGTC CCTNGTTTTT 1980

TTGCAATTTC CTTCTTACTG TGTTAAAAAA AAGTATGATC TTGCTCTGAG AGGTGAGGCA 2040

TTCTTAATCA TGATCTTTAA AGATCAATAA TATAATCCTT TCAAGGATTA TGTCTTTATT 2100

ATAATAAAGA TAATTTGTCT TTAACAGAAT CAATAATATA ATCCCTTAAA GGATTATATC 2160

TTTGCTGGGC GCAGTGGCTC ACACCTGTAA TCCCAGCACT TTGGGTGGCC AAGGTGGAAG 2220

GATCAAATTT GCCTACTTCT ATATTATCTT CTAAAGCAGA ATTCATCTCT CTTCCCTCAA 2280

TATGATGATA TTGACAGGGT TTGCCCTCAC TCACTAGATT GTGAGCTCCT GCTCAGGGCA 2340

GGTAGCGTTT TTTGTTTTTG TTTTTGTTTT TCTTTTTTGA GACAGGGTCT TGCTCTGTCA 2400

CCCAGGCCAG AGTGCAATGG TACAGTCTCA GCTCACTGCA GCCTCAACCG CCTCGGCTCA 2460

AACCATCATC CCATTTCAGC CTCCTGAGTA GCTGGGACTA CAGGCACATG CCATTACACC 2520

TGGCTAATTT TTTTGTATTT CTAGTAGAGA CAGGGTTTGG CCATGTTGCC CGGGCTGGTC 2580

TCGAACTCCT GGACTCAAGC AATCCACCCA CCTCAGCCTC CCAAAATGAG GGACCGTGTC 2640

TTATTCATTT CCATGTCCCT AGTCCATAGC CCAGTGCTGG ACCTATGGTA GTACTAAATA 2700

AATATTTGTT GAATGCAATA GTAAATAGCA TTTCAGGGAG CAAGAACTAG ATTAACAAAG 2760

GTGGTAAAAG GTTTGGAGAA AAAAATAATA GTTTAATTTG GCTAGAGTAT GAGGGAGAGT 2820

AGTAGGAGAC AAGATGGAAA GGTCTCTTGG GCAAGGTTTT GAAGGAAGTT GGAAGTCAGA 2880

AGTACACAAT GTGCATATCG TGGCAGGCAG TGGGGAGCCA ATGAAGGCTT TTGAGCAGGA 2940

GAGTAATGTG TTGAAAAATA AATATAGGTT AAACCTATCA GAGCCCCTCT GACACATACA 3000

CTTGCTTTTC ATTCAAGCTC AAGTTTGTCT CCCACATACC CATTACTTAA CTCACCCTCG 3060

GGCTCCCCTA GCAGCCTGCC CTACCTCTTT ACCTGCTTCC TGGTGGAGTC AGGGATGTAT 3120

ACATGAGCTG CTTTCCCTCT CAGCCAGAGG ACATGGGGGG CCCCAGCTCC CCTGCCTTTC 3180

CCCTTCTGTG CCTGGAGCTG GGAAGCAGGC CAGGGTTAGC TGAGGCTGGC TGGCAAGCAG 3240

CTGGGTGGTG CCAGGGAGAG CCTGCATAGT GCCAGGTGGT GCCTTGGGTT CCAAGCTAGT 3300

CCATGGCCCC GATAACCTTC TGCCTGTGCA CACACCTGCC CCTCACTCCA CCCCCATCCT 3360

AGCTTTGGTA TGGGGGAGAG GGCACAGGGC CAGACAAACC TGTGAGACTT TGGCTCCATC 3420

TCTGCAAAAG GGCGCTCTGT GAGTCAGCCT GCTCCCCTCC AGGCTTGCTC CTCCCCCACC 3480

CAGCTCTCGT TTCCAATGCA CGTACAGCCC GTACACACCG TGTGCTGGGA CACCCCACAG 3540

TCAGCCGCAT GGCTCCCCTG TGCCCCAGCC CCTGGCTCCC TCTGTTGATC CCGGCCCCTG 3600

CTCCAGGCCT CACTGTGCAA CTGCTGCTGT CACTGCTGCT TCTGGTGCCT GTCCATCCCC 3660

AGAGGTTGCC CCGGATGCAG GAGGATTCCC CCTTGGGAGG AGGCTCTTCT GGGGAAGATG 3720

ACCCACTGGG CGAGGAGGAT CTGCCCAGTG AAGAGGATTC ACCCAGAGAG GAGGATCCAC 3780

CCGGAGAGGA GGATCTACCT GGAGAGGAGG ATCTACCTGG AGAGGAGGAT CTACCTGAAG 3840

TTAAGCCTAA ATCAGAAGAA GAGGGCTCCC TGAAGTTAGA GGATCTACCT ACTGTTGAGG 3900

CTCCTGGAGA TCCTCAAGAA CCCCAGAATA ATGCCCACAG GGACAAAGAA GGTAAGTGGT 3960

CATCAATCTC CAAATCCAGG TTCCAGGAGG TTCATGACTC CCCTCCCATA CCCCAGCCTA 4020

GGCTCTGTTC ACTCAGGGAA GGAGGGGAGA CTGTACTCCC CACAGAAGCC CTTCCAGAGG 4080

TCCCATACCA ATATCCCCAT CCCCACTCTC GGAGGTAGAA AGGGACAGAT GTGGAGAGAA 4140

AATAAAAAGG GTGCAAAAGG AGAGAGGTGA GCTGGATGAG ATGGGAGAGA AGGGGGAGGC 4200

TGGAGAAGAG AAAGGGATGA GAACTGCAGA TGAGAGAAAA AATGTGCAGA CAGAGGAAAA 4260

AAATAGGTGG AGAAGGAGAG TCAGAGAGTT TGAGGGGAAG AGAAAAGGAA AGCTTGGGAG 4320

GTGAAGTGGG TACCAGAGAC AAGCAAGAAG AGCTGGTAGA AGTCATCTCA TCTTAGGCTA 4380

CAATGAGGAA TTGAGACCTA GGAAGAAGGG ACACAGCAGG TAGAGAAACG TGGCTTCTTG 4440

ACTCCCAAGC CAGGAATTTG GGGAAAGGGG TTGGAGACCA TACAAGGCAG AGGGATGAGT 4500

GGGGAGAAGA AAGAAGGGAG AAAGGAAAGA TGGTGTACTC ACTCATTTGG GACTCAGGAC 4560

TGAAGTGCCC ACTCACTTTT TTTTTTTTTT TTTTTGAGAC AAACTTTCAC TTTTGTTGCC 4620

CAGGCTGGAG TGCAATGGCG CGATCTCGGC TCACTGCAAC CTCCACCTCC CGGGTTCAAG 4680

TGATTCTCCT GCCTCAGCCT CTAGCCAAGT AGCTGCGATT ACAGGCATGC GCCACCACGC 4740

CCGGCTAATT TTTGTATTTT TAGTAGAGAC GGGGTTTCGC CATGTTGGTC AGGCTGGTCT 4800

CGAACTCCTG ATCTCAGGTG ATCCAACCAC CCTGGCCTCC CAAAGTGCTG GGATTATAGG 4860

CGTGAGCCAC AGCGCCTGGC CTGAAGCAGC CACTCACTTT TACAGACCCT AAGACAATGA 4920

TTGCAAGCTG GTAGGATTGC TGTTTGGCCC ACCCAGCTGC GGTGTTGAGT TTGGGTGCGG 4980

TCTCCTGTGC TTTGCACCTG GCCCGCTTAA GGCATTTGTT ACCCGTAATG CTCCTGTAAG 5040

GCATCTGCGT TTGTGACATC GTTTTGGTCG CCAGGAAGGG ATTGGGGCTC TAAGCTTGAG 5100

CGGTTCATCC TTTTCATTTA TACAGGGGAT GACCAGAGTC ATTGGCGCTA TGGAGGTGAG 5160

ACACCCACCC GCTGCACAGA CCCAATCTGG GAACCCAGCT CTGTGGATCT CCCCTACAGC 5220

CGTCCCTGAA CACTGGTCCC GGGCGTCCCA CCCGCCGCCC ACCGTCCCAC CCCCTCACCT 5280

TTTCTACCCG GGTTCCCTAA GTTCCTGACC TAGGCGTCAG ACTTCCTCAC TATACTCTCC 5340

CACCCCAGGC GACCCGCCCT GGCCCCGGGT GTCCCCAGCC TGCGCGGGCC GCTTCCAGTC 5400

CCCGGTGGAT ATCCGCCCCC AGCTCGCCGC CTTCTGCCCG GCCCTGCGCC CCCTGGAACT 5460

CCTGGGCTTC CAGCTCCCGC CGCTCCCAGA ACTGCGCCTG CGCAACAATG GCCACAGTGG 5520

TGAGGGGGTC TCCCCGCCGA GACTTGGGGA TGGGGCGGGG CGCAGGGAAG GGAACCGTCG 5580

CGCAGTGCCT GCCCGGGGGT TGGGCTGGCC CTACCGGGCG GGGCCGGCTC ACTTGCCTCT 5640

CCCTACGCAG TGCAACTGAC CCTGCCTCCT GGGCTAGAGA TGGCTCTGGG TCCCGGGCGG 5700

GAGTACCGGG CTCTGCAGCT GCATCTGCAC TGGGGGGCTG CAGGTCGTCC GGGCTCGGAG 5760

CACACTGTGG AAGGCCACCG TTTCCCTGCC GAGGTGAGCG CGGACTGGCC GAGAAGGGGC 5820

AAAGGAGCGG GGCGGACGGG GGCCAGAGAC GTGGCCCTCT CCTACCCTCG TGTCCTTTTC 5880

AGATCCACGT GGTTCACCTC AGCACCGCCT TTGCCAGAGT TGACGAGGCC TTGGGGCGCC 5940

CGGGAGGCCT GGCCGTGTTG GCCGCCTTTC TGGAGGTACC AGATCCTGGA CACCCCCTAC 6000

TCCCCGCTTT CCCATCCCAT GCTCCTCCCG GACTCTATCG TGGAGCCAGA GACCCCATCC 6060

CAGCAAGCTC ACTCAGGCCC CTGGCTGACA AACTCATTCA CGCACTGTTT GTTCATTTAA 6120

CACCCACTGT GAACCAGGCA CCAGCCCCCA ACAAGGATTC TGAAGCTGTA GGTCCTTGCC 6180

TCTAAGGAGC CCACAGCCAG TGGGGGAGGC TGACATGACA GACACATAGG AAGGACATAG 6240

TAAAGATGGT GGTCACAGAG GAGGTGACAC TTAAAGCCTT CACTGGTAGA AAAGAAAAGG 6300

AGGTGTTCAT TGCAGAGGAA ACAGAATGTG CAAAGACTCA GAATATGGCC TATTTAGGGA 6360

ATGGCTACAT ACACCATGAT TAGAGGAGGC CCAGTAAAGG GAAGGGATGG TGAGATGCCT 6420

GCTAGGTTCA CTCACTCACT TTTATTTATT TATTTATTTT TTTGACAGTC TCTCTGTCGC 6480

CCAGGCTGGA GTGCAGTGGT GTGATCTTGG GTCACTGCAA CTTCCGCCTC CCGGGTTCAA 6540

GGGATTCTCC TGCCTCAGCT TCCTGAGTAG CTGGGGTTAC AGGTGTGTGC CACCATGCCC 6600

AGCTAATTTT TTTTTGTATT TTTAGTAGAC AGGGTTTCAC CATGTTGGTC AGGCTGGTCT 6660

CAAACTCCTG GCCTCAAGTG ATCCGCCTGA CTCAGCCTAC CAAAGTGCTG ATTACAAGTG 6720

TGAGCCACCG TGCCCAGCCA CACTCACTGA TTCTTTAATG CCAGCCACAC AGCACAAAGT 6780

TCAGAGAAAT GCCTCCATCA TAGCATGTCA ATATGTTCAT ACTCTTAGGT TCATGATGTT 6840

CTTAACATTA GGTTCATAAG CAAAATAAGA AAAAAGAATA ATAAATAAAA GAAGTGGCAT 6900

GTCAGGACCT CACCTGAAAA GCCAAACACA GAATCATGAA GGTGAATGCA GAGGTGACAC 6960

CAACACAAAG GTGTATATAT GGTTTCCTGT GGGGAGTATG TACGGAGGCA GCAGTGAGTG 7020

AGACTGCAAA CGTCAGAAGG GCACGGGTCA CTGAGAGCCT AGTATCCTAG TAAAGTGGGC 7080

TCTCTCCCTC TCTCTCCAGC TTGTCATTGA AAACCAGTCC ACCAAGCTTG TTGGTTCGCA 7140

CAGCAAGAGT ACATAGAGTT TGAAATAATA CATAGGATTT TAAGAGGGAG ACACTGTCTC 7200

TAAAAAAAAA AACAACAGCA ACAACAAAAA GCAACAACCA TTACAATTTT ATGTTCCCTC 7260

AGCATTCTCA GAGCTGAGGA ATGGGAGAGG ACTATGGGAA CCCCCTTCAT GTTCCGGCCT 7320

TCAGCCATGG CCCTGGATAC ATGCACTCAT CTGTCTTACA ATGTCATTCC CCCAGGAGGG 7380

CCCGGAAGAA AACAGTGCCT ATGAGCAGTT GCTGTCTCGC TTGGAAGAAA TCGCTGAGGA 7440

AGGTCAGTTT GTTGGTCTGG CCACTAATCT CTGTGGCCTA GTTCATAAAG AATCACCCTT 7500

TGGAGCTTCA GGTCTGAGGC TGGAGATGGG CTCCCTCCAG TGCAGGAGGG ATTGAAGCAT 7560

GAGCCAGCGC TCATCTTGAT AATAACCATG AAGCTGACAG ACACAGTTAC CCGCAAACGG 7620

CTGCCTACAG ATTGAAAACC AAGCAAAAAC CGCCGGGCAC GGTGGCTCAC GCCTGTAATC 7680

CCAGCACTTT GGGAGGCCAA GGCAGGTGGA TCACGAGGTC AAGAGATCAA GACCATCCTG 7740

GCCAACATGG TGAAACCCCA TCTCTACTAA AAATACGAAA AAATAGCCAG GCGTGGTGGC 7800

GGGTGCCTGT AATCCCAGCT ACTCGGGAGG CTGAGGCAGG AGAATGGCAT GAACCCGGGA 7860

GGCAGAAGTT GCAGTGAGCC GAGATCGTGC CACTGCACTC CAGCCTGGGC AACAGAGCGA 7920

GACTCTTGTC TCAAAAAAAA AAAAAAAAAA GAAAACCAAG CAAAAACCAA AATGAGACAA 7980

AAAAAACAAG ACCAAAAAAT GGTGTTTGGA AATTGTCAAG GTCAAGTCTG GAGAGCTAAA 8040

CTTTTTCTGA GAACTGTTTA TCTTTAATAA GCATCAAATA TTTTAACTTT GTAAATACTT 8100

TTGTTGGAAA TCGTTCTCTT CTTAGTCACT CTTGGGTCAT TTTAAATCTC ACTTACTCTA 8160

CTAGACCTTT TAGGTTTCTG CTAGACTAGG TAGAACTCTG CCTTTGCATT TCTTGTGTCT 8220

GTTTTGTATA GTTATCAATA TTCATATTTA TTTACAAGTT ATTCAGATCA TTTTTTCTTT 8280

TCTTTTTTTT TTTTTTTTTT TTTTTTACAT CTTTAGTAGA GACAGGGTTT CACCATATTG 8340

GCCAGGCTGC TCTCAAACTC CTGACCTTGT GATCCACCAG CCTCGGCCTC CCAAAGTGCT 8400

GGGATTCATT TTTTCTTTTT AATTTGCTCT GGGCTTAAAC TTGTGGCCCA GCACTTTATG 8460

ATGGTACACA GAGTTAAGAG TGTAGACTCA GACGGTCTTT CTTCTTTCCT TCTCTTCCTT 8520

CCTCCCTTCC CTCCCACCTT CCCTTCTCTC CTTCCTTTCT TTCTTCCTCT CTTGCTTCCT 8580

CAGGCCTCTT CCAGTTGCTC CAAAGCCCTG TACTTTTTTT TGAGTTAACG TCTTATGGGA 8640

AGGGCCTGCA CTTAGTGAAG AAGTGGTCTC AGAGTTGAGT TACCTTGGCT TCTGGGAGGT 8700

GAAACTGTAT CCCTATACCC TGAAGCTTTA AGGGGGTGCA ATGTAGATGA GACCCCAACA 8760

TAGATCCTCT TCACAGGCTC AGAGACTCAG GTCCCAGGAC TGGACATATC TGCACTCCTG 8820

CCCTCTGACT TCAGCCGCTA CTTCCAATAT GAGGGGTCTC TGACTACACC GCCCTGTGCC 8880

CAGGGTGTCA TCTGGACTGT GTTTAACCAG ACAGTGATGC TGAGTGCTAA GCAGGTGGGC 8940

CTGGGGTGTG TGTGGACACA GTGGGTGCGG GGGAAAGAGG ATGTAAGATG AGATGAGAAA 9000

CAGGAGAAGA AAGAAATCAA GGCTGGGCTC TGTGGCTTAC GCCTATAATC CCACCACGTT 9060

GGGAGGCTGA GGTGGGAGAA TGGTTTGAGC CCAGGAGTTC AAGACAAGGC GGGGCAACAT 9120

AGTGTGACCC CATCTCTACC AAAAAAACCC CAACAAAACC AAAAATAGCC GGGCATGGTG 9180

GTATGCGGCC TAGTCCCAGC TACTCAAGGA GGCTGAGGTG GGAAGATCGC TTGATTCCAG 9240

GAGTTTGAGA CTGCAGTGAG CTATGATCCC ACCACTGCCT ACCATCTTTA GGATACATTT 9300

ATTTATTTAT AAAAGAAATC AAGAGGCTGG ATGGGGAATA CAGGAGCTGG AGGGTGGAGC 9360

CCTGAGGTGC TGGTTGTGAG CTGGCCTGGG ACCCTTGTTT CCTGTCATGC CATGAACCCA 9420

CCCACACTGT CCACTGACCT CCCTAGCTCC ACACCCTCTC TGACACCCTG TGGGGACCTG 9480

GTGACTCTCG GCTACAGCTG AACTTCCGAG CGACGCAGCC TTTGAATGGG CGAGTGATTG 9540

AGGCCTCCTT CCCTGCTGGA GTGGACAGCA GTCCTCGGGC TGCTGAGCCA GGTACAGCTT 9600

TGTCTGGTTT CCCCCCAGCC AGTAGTCCCT TATCCTCCCA TGTGTGTGCC AGTGTCTGTC 9660

ATTGGTGGTC ACAGCCCGCC TCTCACATCT CCTTTTTCTC TCCAGTCCAG CTGAATTCCT 9720

GCCTGGCTGC TGGTGAGTCT GCCCCTCCTC TTGGTCCTGA TGCCAGGAGA CTCCTCAGCA 9780

CCATTCAGCC CCAGGGCTGC TCAGGACCGC CTCTGCTCCC TCTCCTTTTC TGCAGAACAG 9840

ACCCCAACCC CAATATTAGA GAGGCAGATC ATGGTGGGGA TTCCCCCATT GTCCCCAGAG 9900

GCTAATTGAT TAGAATGAAG CTTGAGAAAT CTCCCAGCAT CCCTCTCGCA AAAGAATCCC 9960

CCCCCCTTTT TTTAAAGATA GGGTCTCACT CTGTTTGCCC CAGGCTGGGG TGTTGTGGCA 10020

CGATCATAGC TCACTGCAGC CTCGAACTCC TAGGCTCAGG CAATCCTTTC ACCTTAGCTT 10080

CTCAAAGCAC TGGGACTGTA GGCATGAGCC ACTGTGCCTG GCCCCAAACG GCCCTTTTAC 10140

TTGGCTTTTA GGAAGCAAAA ACGGTGCTTA TCTTACCCCT TCTCGTGTAT CCACCCTCAT 10200

CCCTTGGCTG GCCTCTTCTG GAGACTGAGG CACTATGGGG CTGCCTGAGA ACTCGGGGCA 10260

GGGGTGGTGG AGTGCACTGA GGCAGGTGTT GAGGAACTCT GCAGACCCCT CTTCCTTCCC 10320

AAAGCAGCCC TCTCTGCTCT CCATCGCAGG TGACATCCTA GCCCTGGTTT TTGGCCTCCT 10380

TTTTGCTGTC ACCAGCGTCG CGTTCCTTGT GCAGATGAGA AGGCAGCACA GGTATTACAC 10440

TGACCCTTTC TTCAGGCACA AGCTTCCCCC ACCCTTGTGG AGTCACTTCA TGCAAAGCGC 10500

ATGCAAATGA GCTGCTCCTG GGCCAGTTTT CTGATTAGCC TTTCCTGTTG TGTACACACA 10560

GAAGGGGAAC CAAAGGGGGT GTGAGCTACC GCCCAGCAGA GGTAGCCGAG ACTGGAGCCT 10620

AGAGGCTGGA TCTTGGAGAA TGTGAGAAGC CAGCCAGAGG CATCTGAGGG GGAGCCGGTA 10680

ACTGTCCTGT CCTGCTCATT ATGCCACTTC CTTTTAACTG CCAAGAAATT TTTTAAAATA 10740

AATATTTATA ATAAAATATG TGTTAGTCAC CTTTGTTCCC CAAATCAGAA GGAGGTATTT 10800

GAATTTCCTA TTACTGTTAT TAGCACCAAT TTAGTGGTAA TGCATTTATT CTATTACAGT 10860

TCGGCCTCCT TCCACACATC ACTCCAATGT GTTGCTCC 10898

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: Signal peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Ala Pro Leu Cys Pro Ser Pro Trp Leu Pro Leu Leu Ile Pro Ala
1 5 10 15

Pro Ala Pro Gly Leu Thr Val Gln Leu Leu Leu Ser Leu Leu Leu Leu
20 25 30

Met Pro Val His Pro
35

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TGGGGTTCTT GAGGATCTCC AGGAG 25

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CTCTAACTTC AGGGAGCCCT CTTCTT 26

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

(D) OTHER INFORMATION: N stands for inosine

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CUACUACUAC UAGGCCACGC GTCGACTAGT ACGGGNNGGG NNGGGNNG 48

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Glu Glu Asp Leu Pro Ser
1 5

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION:55..60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Gly Glu Asp Asp Pro Leu
1 5

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Asn Asn Ala His Arg Asp Lys Glu Gly Asp Asp Gln Ser His Trp Arg
1 5 10 15

Tyr Gly Gly Asp Pro
20

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION:36..51

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

His Pro Gln Arg Leu Pro Arg Met Gln Glu Asp Ser Pro Leu Gly Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Glu Glu Asp Ser Pro Arg Glu Glu Asp Pro Pro Gly Glu Glu Asp Leu
1 5 10 15

Pro Gly Glu Glu Asp Leu Pro Gly
20

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION:279..291

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Leu Glu Glu Gly Pro Glu Glu Asn Ser Ala Tyr Glu Gln
1 5 10

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Arg Arg Gln His Arg Arg Gly Thr Lys Gly Gly Val Ser Tyr Arg
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GTCGCTAGCT CCATGGGTCA TATGCAGAGG TTGCCCCGGA TGCAG 45

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GAAGATCTCT TACTCGAGCA TTCTCCAAGA TCCAGCCTCT AGG 43

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: DNA (genomic)
(A) DESCRIPTION: AP-2 transcription factor

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TCCCCCACCC 10

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: DNA (genomic)
(A) DESCRIPTION: initiator (Inr) element

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCACCCCCAT 10

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: DNA (genomic)
(A) DESCRIPTION: p53 binding site

(x) PUBLICATION INFORMATION:
(A) AUTHORS: El Deiry et al.
(B) TITLE: "Human genomic DNA sequences define a consensus binding site for p53"
(C) JOURNAL: Nature Genetics
(D) VOLUME: 1
(F) PAGES: 44-49
(G) DATE: 1992

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AAGCTAGTCC 10

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

Leu Glu His His His His His His
1 5

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: Initiator consensus sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

YYYCAYYYYY 10

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)
(A) DESCRIPTION: p53 binding site

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(x) PUBLICATION INFORMATION:
(A) AUTHORS: El Deiry et al.
(B) TITLE: "Human genomic DNA sequences define a consensus binding site for p53"
(C) JOURNAL: Nature Genetics
(D) VOLUME: 1
(F) PAGES: 44-49
(G) DATE: 1992

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AGGCTTGCTC 10

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

Ser Pro Xaa Xaa
1

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Thr Pro Xaa Xaa
1

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 540 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: Proposed MN promoter

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CTTGCTTTTC ATTCAAGCTC AAGTTTGTCT CCCACATACC CATTACTTAA CTCACCCTCG 60

GGCTCCCCTA GCAGCCTGCC CTACCTCTTT ACCTGCTTCC TGGTGGAGTC AGGGATGTAT 120

ACATGAGCTG CTTTCCCTCT CAGCCAGAGG ACATGGGGGG CCCCAGCTCC CCTGCCTTTC 180

CCCTTCTGTG CCTGGAGCTG GGAAGCAGGC CAGGGTTAGC TGAGGCTGGC TGGCAAGCAG 240

CTGGGTGGTG CCAGGGAGAG CCTGCATAGT GCCAGGTGGT GCCTTGGGTT CCAAGCTAGT 300

CCATGGCCCC GATAACCTTC TGCCTGTGCA CACACCTGCC CCTCACTCCA CCCCCATCCT 360

AGCTTTGGTA TGGGGGAGAG GGCACAGGGC CAGACAAACC TGTGAGACTT TGGCTCCATC 420

TCTGCAAAAG GGCGCTCTGT GAGTCAGCCT GCTCCCCTCC AGGCTTGCTC CTCCCCCACC 480

CAGCTCTCGT TTCCAATGCA CGTACAGCCC GTACACACCG TGTGCTGGGA CACCCCACAG 540

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 415 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 1st MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

ACAGTCAGCC GCATGGCTCC CCTGTGCCCC AGCCCCTGGC TCCCTCTGTT GATCCCGGCC 60

CCTGCTCCAG GCCTCACTGT GCAACTGCTG CTGTCACTGC TGCTTCTGGT GCCTGTCCAT 120

CCCCAGAGGT TGCCCCGGAT GCAGGAGGAT TCCCCCTTGG GAGGAGGCTC TTCTGGGGAA 180

GATGACCCAC TGGGCGAGGA GGATCTGCCC AGTGAAGAGG ATTCACCCAG AGAGGAGGAT 240

CCACCCGGAG AGGAGGATCT ACCTGGAGAG GAGGATCTAC CTGGAGAGGA GGATCTACCT 300

GAAGTTAAGC CTAAATCAGA AGAAGAGGGC TCCCTGAAGT TAGAGGATCT ACCTACTGTT 360

GAGGCTCCTG GAGATCCTCA AGAACCCCAG AATAATGCCC ACAGGGACAA AGAAG 415

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 2nd MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GGGATGACCA GAGTCATTGG CGCTATGGAG 30

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 171 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3rd MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

GCGACCCGCC CTGGCCCCGG GTGTCCCCAG CCTGCGCGGG CCGCTTCCAG TCCCCGGTGG 60

ATATCCGCCC CCAGCTCGCC GCCTTCTGCC CGGCCCTGCG CCCCCTGGAA CTCCTGGGCT 120

TCCAGCTCCC GCCGCTCCCA GAACTGCGCC TGCGCAACAA TGGCCACAGT G 171

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 143 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 4th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

TGCAACTGAC CCTGCCTCCT GGGCTAGAGA TGGCTCTGGG TCCCGGGCGG GAGTACCGGG 60

CTCTGCAGCT GCATCTGCAC TGGGGGGCTG CAGGTCGTCC GGGCTCGGAG CACACTGTGG 120

AAGGCCACCG TTTCCCTGCC GAG 143

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 93 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

ATCCACGTGG TTCACCTCAG CACCGCCTTT GCCAGAGTTG ACGAGGCCTT GGGGCGCCCG 60

GGAGGCCTGG CCGTGTTGGC CGCCTTTCTG GAG 93

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 67 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 6th MN exon

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

GAGGGCCCGG AAGAAAACAG TGCCTATGAG CAGTTGCTGT CTCGCTTGGA AGAAATCGCT 60

GAGGAAG 67

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 158 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 7th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GCTCAGAGAC TCAGGTCCCA GGACTGGACA TATCTGCACT CCTGCCCTCT GACTTCAGCC 60

GCTACTTCCA ATATGAGGGG TCTCTGACTA CACCGCCCTG TGCCCAGGGT GTCATCTGGA 120

CTGTGTTTAA CCAGACAGTG ATGCTGAGTG CTAAGCAG 158

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 145 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 8th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CTCCACACCC TCTCTGACAC CCTGTGGGGA CCTGGTGACT CTCGGCTACA GCTGAACTTC 60

CGAGCGACGC AGCCTTTGAA TGGGCGAGTG ATTGAGGCCT CCTTCCCTGC TGGAGTGGAC 120

AGCAGTCCTC GGGCTGCTGA GCCAG 145

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 9th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

TCCAGCTGAA TTCCTGCCTG GCTGCTG 27

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 82 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 10th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GTGACATCCT AGCCCTGGTT TTTGGCCTCC TTTTTGCTGT CACCAGCGTC GCGTTCCTTG 60

TGCAGATGAG AAGGCAGCAC AG 82

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 191 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 11th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

AAGGGGAACC AAAGGGGGTG TGAGCTACCG CCCAGCAGAG GTAGCCGAGA CTGGAGCCTA 60

GAGGCTGGAT CTTGGAGAAT GTGAGAAGCC AGCCAGAGGC ATCTGAGGGG GAGCCGGTAA 120

CTGTCCTGTC CTGCTCATTA TGCCACTTCC TTTTAACTGC CAAGAAATTT TTTAAAATAA 180

ATATTTATAA T 191

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1174 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 1st MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

GTAAGTGGTC ATCAATCTCC AAATCCAGGT TCCAGGAGGT TCATGACTCC CCTCCCATAC 60

CCCAGCCTAG GCTCTGTTCA CTCAGGGAAG GAGGGGAGAC TGTACTCCCC ACAGAAGCCC 120

TTCCAGAGGT CCCATACCAA TATCCCCATC CCCACTCTCG GAGGTAGAAA GGGACAGATG 180

TGGAGAGAAA ATAAAAAGGG TGCAAAAGGA GAGAGGTGAG CTGGATGAGA TGGGAGAGAA 240

GGGGGAGGCT GGAGAAGAGA AAGGGATGAG AACTGCAGAT GAGAGAAAAA ATGTGCAGAC 300

AGAGGAAAAA AATAGGTGGA GAAGGAGAGT CAGAGAGTTT GAGGGGAAGA GAAAAGGAAA 360

GCTTGGGAGG TGAAGTGGGT ACCAGAGACA AGCAAGAAGA GCTGGTAGAA GTCATCTCAT 420

CTTAGGCTAC AATGAGGAAT TGAGACCTAG GAAGAAGGGA CACAGCAGGT AGAGAAACGT 480

GGCTTCTTGA CTCCCAAGCC AGGAATTTGG GGAAAGGGGT TGGAGACCAT ACAAGGCAGA 540

GGGATGAGTG GGGAGAAGAA AGAAGGGAGA AAGGAAAGAT GGTGTACTCA CTCATTTGGG 600

ACTCAGGACT GAAGTGCCCA CTCACTTTTT TTTTTTTTTT TTTTGAGACA AACTTTCACT 660

TTTGTTGCCC AGGCTGGAGT GCAATGGCGC GATCTCGGCT CACTGCAACC TCCACCTCCC 720

GGGTTCAAGT GATTCTCCTG CCTCAGCCTC TAGCCAAGTA GCTGCGATTA CAGGCATGCG 780

CCACCACGCC CGGCTAATTT TTGTATTTTT AGTAGAGACG GGGTTTCGCC ATGTTGGTCA 840

GGCTGGTCTC GAACTCCTGA TCTCAGGTGA TCCAACCACC CTGGCCTCCC AAAGTGCTGG 900

GATTATAGGC GTGAGCCACA GCGCCTGGCC TGAAGCAGCC ACTCACTTTT ACAGACCCTA 960

AGACAATGAT TGCAAGCTGG TAGGATTGCT GTTTGGCCCA CCCAGCTGCG GTGTTGAGTT 1020

TGGGTGCGGT CTCCTGTGCT TTGCACCTGG CCCGCTTAAG GCATTTGTTA CCCGTAATGC 1080

TCCTGTAAGG CATCTGCGTT TGTGACATCG TTTTGGTCGC CAGGAAGGGA TTGGGGCTCT 1140

AAGCTTGAGC GGTTCATCCT TTTCATTTAT ACAG 1174

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 193 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 2nd MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

GTGAGACACC CACCCGCTGC ACAGACCCAA TCTGGGAACC CAGCTCTGTG GATCTCCCCT 60

ACAGCCGTCC CTGAACACTG GTCCCGGGCG TCCCACCCGC CGCCCACCGT CCCACCCCCT 120

CACCTTTTCT ACCCGGGTTC CCTAAGTTCC TGACCTAGGC GTCAGACTTC CTCACTATAC 180

TCTCCCACCC CAG 193

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 131 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3rd MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GTGAGGGGGT CTCCCCGCCG AGACTTGGGG ATGGGGCGGG GCGCAGGGAA GGGAACCGTC 60

GCGCAGTGCC TGCCCGGGGG TTGGGCTGGC CCTACCGGGC GGGGCCGGCT CACTTGCCTC 120

TCCCTACGCA G 131

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 89 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 4th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

GTGAGCGCGG ACTGGCCGAG AAGGGGCAAA GGAGCGGGGC GGACGGGGGC CAGAGACGTG 60

GCCCTCTCCT ACCCTCGTGT CCTTTTCAG 89

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1400 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

GTACCAGATC CTGGACACCC CCTACTCCCC GCTTTCCCAT CCCATGCTCC TCCCGGACTC 60

TATCGTGGAG CCAGAGACCC CATCCCAGCA AGCTCACTCA GGCCCCTGGC TGACAAACTC 120

ATTCACGCAC TGTTTGTTCA TTTAACACCC ACTGTGAACC AGGCACCAGC CCCCAACAAG 180

GATTCTGAAG CTGTAGGTCC TTGCCTCTAA GGAGCCCACA GCCAGTGGGG GAGGCTGACA 240

TGACAGACAC ATAGGAAGGA CATAGTAAAG ATGGTGGTCA CAGAGGAGGT GACACTTAAA 300

GCCTTCACTG GTAGAAAAGA AAAGGAGGTG TTCATTGCAG AGGAAACAGA ATGTGCAAAG 360

ACTCAGAATA TGGCCTATTT AGGGAATGGC TACATACACC ATGATTAGAG GAGGCCCAGT 420

AAAGGGAAGG GATGGTGAGA TGCCTGCTAG GTTCACTCAC TCACTTTTAT TTATTTATTT 480

ATTTTTTTGA CAGTCTCTCT GTCGCCCAGG CTGGAGTGCA GTGGTGTGAT CTTGGGTCAC 540

TGCAACTTCC GCCTCCCGGG TTCAAGGGAT TCTCCTGCCT CAGCTTCCTG AGTAGCTGGG 600

GTTACAGGTG TGTGCCACCA TGCCCAGCTA ATTTTTTTTT GTATTTTTAG TAGACAGGGT 660

TTCACCATGT TGGTCAGGCT GGTCTCAAAC TCCTGGCCTC AAGTGATCCG CCTGACTCAG 720

CCTACCAAAG TGCTGATTAC AAGTGTGAGC CACCGTGCCC AGCCACACTC ACTGATTCTT 780

TAATGCCAGC CACACAGCAC AAAGTTCAGA GAAATGCCTC CATCATAGCA TGTCAATATG 840

TTCATACTCT TAGGTTCATG ATGTTCTTAA CATTAGGTTC ATAAGCAAAA TAAGAAAAAA 900

GAATAATAAA TAAAAGAAGT GGCATGTCAG GACCTCACCT GAAAAGCCAA ACACAGAATC 960

ATGAAGGTGA ATGCAGAGGT GACACCAACA CAAAGGTGTA TATATGGTTT CCTGTGGGGA 1020

GTATGTACGG AGGCAGCAGT GAGTGAGACT GCAAACGTCA GAAGGGCACG GGTCACTGAG 1080

AGCCTAGTAT CCTAGTAAAG TGGGCTCTCT CCCTCTCTCT CCAGCTTGTC ATTGAAAACC 1140

AGTCCACCAA GCTTGTTGGT TCGCACAGCA AGAGTACATA GAGTTTGAAA TAATACATAG 1200

GATTTTAAGA GGGAGACACT GTCTCTAAAA AAAAAAACAA CAGCAACAAC AAAAAGCAAC 1260

AACCATTACA ATTTTATGTT CCCTCAGCAT TCTCAGAGCT GAGGAATGGG AGAGGACTAT 1320

GGGAACCCCC TTCATGTTCC GGCCTTCAGC CATGGCCCTG GATACATGCA CTCATCTGTC 1380

TTACAATGTC ATTCCCCCAG 1400

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1334 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 6th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GTCAGTTTGT TGGTCTGGCC ACTAATCTCT GTGGCCTAGT TCATAAAGAA TCACCCTTTG 60

GAGCTTCAGG TCTGAGGCTG GAGATGGGCT CCCTCCAGTG CAGGAGGGAT TGAAGCATGA 120

GCCAGCGCTC ATCTTGATAA TAACCATGAA GCTGACAGAC ACAGTTACCC GCAAACGGCT 180

GCCTACAGAT TGAAAACCAA GCAAAAACCG CCGGGCACGG TGGCTCACGC CTGTAATCCC 240

AGCACTTTGG GAGGCCAAGG CAGGTGGATC ACGAGGTCAA GAGATCAAGA CCATCCTGGC 300

CAACATGGTG AAACCCCATC TCTACTAAAA ATACGAAAAA ATAGCCAGGC GTGGTGGCGG 360

GTGCCTGTAA TCCCAGCTAC TCGGGAGGCT GAGGCAGGAG AATGGCATGA ACCCGGGAGG 420

CAGAAGTTGC AGTGAGCCGA GATCGTGCCA CTGCACTCCA GCCTGGGCAA CAGAGCGAGA 480

CTCTTGTCTC AAAAAAAAAA AAAAAAAAGA AAACCAAGCA AAAACCAAAA TGAGACAAAA 540

AAAACAAGAC CAAAAAATGG TGTTTGGAAA TTGTCAAGGT CAAGTCTGGA GAGCTAAACT 600

TTTTCTGAGA ACTGTTTATC TTTAATAAGC ATCAAATATT TTAACTTTGT AAATACTTTT 660

GTTGGAAATC GTTCTCTTCT TAGTCACTCT TGGGTCATTT TAAATCTCAC TTACTCTACT 720

AGACCTTTTA GGTTTCTGCT AGACTAGGTA GAACTCTGCC TTTGCATTTC TTGTGTCTGT 780

TTTGTATAGT TATCAATATT CATATTTATT TACAAGTTAT TCAGATCATT TTTTCTTTTC 840

TTTTTTTTTT TTTTTTTTTT TTTTACATCT TTAGTAGAGA CAGGGTTTCA CCATATTGGC 900

CAGGCTGCTC TCAAACTCCT GACCTTGTGA TCCACCAGCC TCGGCCTCCC AAAGTGCTGG 960

GATTCATTTT TTCTTTTTAA TTTGCTCTGG GCTTAAACTT GTGGCCCAGC ACTTTATGAT 1020

GGTACACAGA GTTAAGAGTG TAGACTCAGA CGGTCTTTCT TCTTTCCTTC TCTTCCTTCC 1080

TCCCTTCCCT CCCACCTTCC CTTCTCTCCT TCCTTTCTTT CTTCCTCTCT TGCTTCCTCA 1140

GGCCTCTTCC AGTTGCTCCA AAGCCCTGTA CTTTTTTTTG AGTTAACGTC TTATGGGAAG 1200

GGCCTGCACT TAGTGAAGAA GTGGTCTCAG AGTTGAGTTA CCTTGGCTTC TGGGAGGTGA 1260

AACTGTATCC CTATACCCTG AAGCTTTAAG GGGGTGCAAT GTAGATGAGA CCCCAACATA 1320

GATCCTCTTC ACAG 1334

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 512 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 7th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GTGGGCCTGG GGTGTGTGTG GACACAGTGG GTGCGGGGGA AAGAGGATGT AAGATGAGAT 60

GAGAAACAGG AGAAGAAAGA AATCAAGGCT GGGCTCTGTG GCTTACGCCT ATAATCCCAC 120

CACGTTGGGA GGCTGAGGTG GGAGAATGGT TTGAGCCCAG GAGTTCAAGA CAAGGCGGGG 180

CAACATAGTG TGACCCCATC TCTACCAAAA AAACCCCAAC AAAACCAAAA ATAGCCGGGC 240

ATGGTGGTAT GCGGCCTAGT CCCAGCTACT CAAGGAGGCT GAGGTGGGAA GATCGCTTGA 300

TTCCAGGAGT TTGAGACTGC AGTGAGCTAT GATCCCACCA CTGCCTACCA TCTTTAGGAT 360

ACATTTATTT ATTTATAAAA GAAATCAAGA GGCTGGATGG GGAATACAGG AGCTGGAGGG 420

TGGAGCCCTG AGGTGCTGGT TGTGAGCTGG CCTGGGACCC TTGTTTCCTG TCATGCCATG 480

AACCCACCCA CACTGTCCAC TGACCTCCCT AG 512

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 114 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 8th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

GTACAGCTTT GTCTGGTTTC CCCCCAGCCA GTAGTCCCTT ATCCTCCCAT GTGTGTGCCA 60

GTGTCTGTCA TTGGTGGTCA CAGCCCGCCT CTCACATCTC CTTTTTCTCT CCAG 114

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 617 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 9th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

GTGAGTCTGC CCCTCCTCTT GGTCCTGATG CCAGGAGACT CCTCAGCACC ATTCAGCCCC 60

AGGGCTGCTC AGGACCGCCT CTGCTCCCTC TCCTTTTCTG CAGAACAGAC CCCAACCCCA 120

ATATTAGAGA GGCAGATCAT GGTGGGGATT CCCCCATTGT CCCCAGAGGC TAATTGATTA 180

GAATGAAGCT TGAGAAATCT CCCAGCATCC CTCTCGCAAA AGAATCCCCC CCCCTTTTTT 240

TAAAGATAGG GTCTCACTCT GTTTGCCCCA GGCTGGGGTG TTGTGGCACG ATCATAGCTC 300

ACTGCAGCCT CGAACTCCTA GGCTCAGGCA ATCCTTTCAC CTTAGCTTCT CAAAGCACTG 360

GGACTGTAGG CATGAGCCAC TGTGCCTGGC CCCAAACGGC CCTTTTACTT GGCTTTTAGG 420

AAGCAAAAAC GGTGCTTATC TTACCCCTTC TCGTGTATCC ACCCTCATCC CTTGGCTGGC 480

CTCTTCTGGA GACTGAGGCA CTATGGGGCT GCCTGAGAAC TCGGGGCAGG GGTGGTGGAG 540

TGCACTGAGG CAGGTGTTGA GGAACTCTGC AGACCCCTCT TCCTTCCCAA AGCAGCCCTC 600

TCTGCTCTCC ATCGCAG 617

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 130 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 10th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

GTATTACACT GACCCTTTCT TCAGGCACAA GCTTCCCCCA CCCTTGTGGA GTCACTTCAT 60

GCAAAGCGCA TGCAAATGAG CTGCTCCTGG GCCAGTTTTC TGATTAGCCT TTCCTGTTGT 120

GTACACACAG 130

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1401 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: Spans 3' part of 1st intron to beyond end of 5th exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

CAAACTTTCA CTTTTGTTGC CCAGGCTGGA GTGCAATGGC GCGATCTCGG CTCACTGCAA 60

CCTCCACCTC CCGGGTTCAA GTGATTCTCC TGCCTCAGCC TCTAGCCAAG TAGCTGCGAT 120

TACAGGCATG CGCCACCACG CCCGGCTAAT TTTTGTATTT TTAGTAGAGA CGGGGTTTCG 180

CCATGTTGGT CAGGCTGGTC TCGAACTCCT GATCTCAGGT GATCCAACCA CCCTGGCCTC 240

CCAAAGTGCT GGGATTATAG GCGTGAGCCA CAGCGCCTGG CCTGAAGCAG CCACTCACTT 300

TTACAGACCC TAAGACAATG ATTGCAAGCT GGTAGGATTG CTGTTTGGCC CACCCAGCTG 360

CGGTGTTGAG TTTGGGTGCG GTCTCCTGTG CTTTGCACCT GGCCCGCTTA AGGCATTTGT 420

TACCCGTAAT GCTCCTGTAA GGCATCTGCG TTTGTGACAT CGTTTTGGTC GCCAGGAAGG 480

GATTGGGGCT CTAAGCTTGA GCGGTTCATC CTTTTCATTT ATACAGGGGA TGACCAGAGT 540

CATTGGCGCT ATGGAGGTGA GACACCCACC CGCTGCACAG ACCCAATCTG GGAACCCAGC 600

TCTGTGGATC TCCCCTACAG CCGTCCCTGA ACACTGGTCC CGGGCGTCCC ACCCGCCGCC 660

CACCGTCCCA CCCCCTCACC TTTTCTACCC GGGTTCCCTA AGTTCCTGAC CTAGGCGTCA 720

GACTTCCTCA CTATACTCTC CCACCCCAGG CGACCCGCCC TGGCCCCGGG TGTCCCCAGC 780

CTGCGCGGGC CGCTTCCAGT CCCCGGTGGA TATCCGCCCC CAGCTCGCCG CCTTCTGCCC 840

GGCCCTGCGC CCCCTGGAAC TCCTGGGCTT CCAGCTCCCG CCGCTCCCAG AACTGCGCCT 900

GCGCAACAAT GGCCACAGTG GTGAGGGGGT CTCCCCGCCG AGACTTGGGG ATGGGGCGGG 960

GCGCAGGGAA GGGAACCGTC GCGCAGTGCC TGCCCGGGGG TTGGGCTGGC CCTACCGGGC 1020

GGGGCCGGCT CACTTGCCTC TCCCTACGCA GTGCAACTGA CCCTGCCTCC TGGGCTAGAG 1080

ATGGCTCTGG GTCCCGGGCG GGAGTACCGG GCTCTGCAGC TGCATCTGCA CTGGGGGGCT 1140

GCAGGTCGTC CGGGCTCGGA GCACACTGTG GAAGGCCACC GTTTCCCTGC CGAGGTGAGC 1200

GCGGACTGGC CGAGAAGGGG CAAAGGAGCG GGGCGGACGG GGGCCAGAGA CGTGGCCCTC 1260

TCCTACCCTC GTGTCCTTTT CAGATCCACG TGGTTCACCT CAGCACCGCC TTTGCCAGAG 1320

TTGACGAGGC CTTGGGGCGC CCGGGAGGCC TGGCCGTGTT GGCCGCCTTT CTGGAGGTAC 1380

CAGATCCTGG ACACCCCCTA C 1401

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 98 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(A) DESCRIPTION: Region of homology to collagen alpha 1 chain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

```
Gln Arg Leu Pro Arg Met Gln Glu Asp Ser Pro Leu Gly Gly Gly Ser
1               5                  10                  15

Ser Gly Glu Asp Asp Pro Leu Gly Glu Glu Asp Leu Pro Ser Glu Glu
            20                  25                  30

Asp Ser Pro Arg Glu Glu Asp Pro Pro Gly Glu Glu Asp Leu Pro Gly
        35                  40                  45

Glu Glu Asp Leu Pro Gly Glu Glu Asp Leu Pro Glu Val Lys Pro Lys
    50                  55                  60

Ser Glu Glu Glu Gly Ser Leu Lys Leu Glu Asp Leu Pro Thr Val Glu
65                  70                  75                  80

Ala Pro Gly Asp Pro Gln Glu Pro Gln Asn Asn Ala His Arg Asp Lys
                85                  90                  95

Glu Gly
```

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 256 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(A) DESCRIPTION: carbonic anhydrase domain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

```
Asp Asp Gln Ser His Trp Arg Tyr Gly Gly Asp Pro Pro Trp Pro Arg
1               5                  10                  15

Val Ser Pro Ala Cys Ala Gly Arg Phe Gln Ser Pro Val Asp Ile Arg
            20                  25                  30

Pro Gln Leu Ala Ala Phe Cys Pro Ala Leu Arg Pro Leu Glu Leu Leu
        35                  40                  45

Gly Phe Gln Leu Pro Pro Leu Pro Glu Leu Arg Leu Arg Asn Asn Gly
    50                  55                  60

His Ser Val Gln Leu Thr Leu Pro Pro Gly Leu Glu Met Ala Leu Gly
65                  70                  75                  80

Pro Gly Arg Glu Tyr Arg Ala Leu Gln Leu His Leu His Trp Gly Ala
                85                  90                  95

Ala Gly Arg Pro Gly Ser Glu His Thr Val Glu Gly His Arg Phe Pro
            100                 105                 110

Ala Glu Ile His Val Val His Leu Ser Thr Ala Phe Ala Arg Val Asp
        115                 120                 125

Glu Ala Leu Gly Arg Pro Gly Gly Leu Ala Val Leu Ala Ala Phe Leu
    130                 135                 140

Glu Glu Gly Pro Glu Glu Asn Ser Ala Tyr Glu Gln Leu Leu Ser Arg
145                 150                 155                 160

Leu Glu Glu Ile Ala Glu Glu Gly Ser Glu Thr Gln Val Pro Gly Leu
                165                 170                 175
```

```
Asp Ile Ser Ala Leu Leu Pro Ser Asp Phe Ser Arg Tyr Phe Gln Tyr
                180                 185                 190

Glu Gly Ser Leu Thr Thr Pro Pro Cys Ala Gln Gly Val Ile Trp Thr
        195                 200                 205

Val Phe Asn Gln Thr Val Met Leu Ser Ala Lys Gln Leu His Thr Leu
    210                 215                 220

Ser Asp Thr Leu Trp Gly Pro Gly Asp Ser Arg Leu Gln Leu Asn Phe
225                 230                 235                 240

Arg Ala Thr Gln Pro Leu Asn Gly Arg Val Ile Glu Ala Ser Phe Pro
                245                 250                 255
```

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: transmembrane region

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

```
Asp Ile Leu Ala Leu Val Phe Gly Leu Leu Phe Ala Val Thr Ser Val
1               5                   10                  15

Ala Phe Leu Val
            20
```

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: intracellular C-terminus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

```
Met Arg Arg Gln His Arg Arg Gly Thr Lys Gly Gly Val Ser Tyr Arg
1               5                  10                  15

Pro Ala Glu Val Ala Glu Thr Gly Ala
            20                  25
```

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 170 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

```
Arg Ala Leu Gln Leu His Leu His Trp Gly Ala Ala Gly Arg Pro Gly
1               5                  10                  15

Ser Glu His Thr Val Glu Gly His Arg Phe Pro Ala Glu Ile His Val
            20                  25                  30

Val His Leu Ser Thr Ala Phe Ala Arg Val Asp Glu Ala Leu Gly Arg
        35                  40                  45

Pro Gly Gly Leu Ala Val Leu Ala Ala Phe Leu Glu Glu Gly Pro Glu
    50                  55                  60

Glu Asn Ser Ala Tyr Glu Gln Leu Leu Ser Arg Leu Glu Glu Ile Ala
65                  70                  75                  80

Glu Glu Gly Ser Glu Thr Gln Val Pro Gly Leu Asp Ile Ser Ala Leu
                85                  90                  95
```

Leu Pro Ser Asp Phe Ser Arg Tyr Phe Gln Tyr Glu Gly Ser Leu Thr
100 105 110

Thr Pro Pro Cys Ala Gln Gly Val Ile Trp Thr Val Phe Asn Gln Thr
115 120 125

Val Met Leu Ser Ala Lys Gln Leu His Thr Leu Ser Asp Thr Leu Trp
130 135 140

Gly Pro Gly Asp Ser Arg Leu Gln Leu Asn Phe Arg Ala Thr Gln Pro
145 150 155 160

Leu Asn Gly Arg Val Ile Glu Ala Ser Phe
165 170

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 470 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

CAUGGCCCCG AUAACCUUCU GCCUGUGCAC ACACCUGCCC CUCACUCCAC CCCCAUCCUA 60

GCUUUGGUAU GGGGGAGAGG GCACAGGGCC AGACAAACCU GUGAGACUUU GGCUCCAUCU 120

CUGCAAAAGG GCGCUCUGUG AGUCAGCCUG CUCCCCUCCA GGCUUGCUCC UCCCCCACCC 180

AGCUCUCGUU UCCAAUGCAC GUACAGCCCG UACACACCGU GUGCUGGGAC ACCCCACAGU 240

CAGCCGCAUG GCUCCCCUGU GCCCCAGCCC CUGGCUCCCU CUGUUGAUCC CGGCCCCUGC 300

UCCAGGCCUC ACUGUGCAAC UGCUGCUGUC ACUGCUGCUU CUGGUGCCUG UCCAUCCCCA 360

GAGGUUGCCC CGGAUGCAGG AGGAUUCCCC CUUGGGAGGA GGCUCUUCUG GGGAAGAUGA 420

CCCACUGGGC GAGGAGGAUC UGCCCAGUGA AGAGGAUUCA CCCAGAGAGG 470

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

This sequence is intentionally skipped.

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

This sequence is intentionally skipped.

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 904 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

GCTGGTCTCG AACTCCTGGA CTCAAGCAAT CCACCCACCT CAGCCTCCCA AAATGAGGGA 60

CCGTGTCTTA TTCATTTCCA TGTCCCTAGT CCATAGCCCA GTGCTGGACC TATGGTAGTA 120

CTAAATAAAT ATTTGTTGAA TGCAATAGTA AATAGCATTT CAGGGAGCAA GAACTAGATT 180

AACAAAGGTG GTAAAAGGTT TGGAGAAAAA AATAATAGTT TAATTTGGCT AGAGTATGAG 240

GGAGAGTAGT AGGAGACAAG ATGGAAAGGT CTCTTGGGCA AGGTTTTGAA GGAAGTTGGA 300

AGTCAGAAGT ACACAATGTG CATATCGTGG CAGGCAGTGG GGAGCCAATG AAGGCTTTTG 360

AGCAGGAGAG TAATGTGTTG AAAAATAAAT ATAGGTTAAA CCTATCAGAG CCCCTCTGAC 420

ACATACACTT GCTTTTCATT CAAGCTCAAG TTTGTCTCCC ACATACCCAT TACTTAACTC 480

ACCCTCGGGC TCCCCTAGCA GCCTGCCCTA CCTCTTTACC TGCTTCCTGG TGGAGTCAGG 540

GATGTATACA TGAGCTGCTT TCCCTCTCAG CCAGAGGACA TGGGGGGCCC CAGCTCCCCT 600

GCCTTTCCCC TTCTGTGCCT GGAGCTGGGA AGCAGGCCAG GGTTAGCTGA GGCTGGCTGG 660

CAAGCAGCTG GGTGGTGCCA GGGAGAGCCT GCATAGTGCC AGGTGGTGCC TTGGGTTCCA 720

AGCTAGTCCA TGGCCCCGAT AACCTTCTGC CTGTGCACAC ACCTGCCCCT CACTCCACCC 780

CCATCCTAGC TTTGGTATGG GGGAGAGGGC ACAGGGCCAG ACAAACCTGT GAGACTTTGG 840

CTCCATCTCT GCAAAAGGGC GCTCTGTGAG TCAGCCTGCT CCCCTCCAGG CTTGCTCCTC 900

CCCC 904

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 292 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

TTTTTTTGAG ACGGAGTCTT GCATCTGTCA TGCCCAGGCT GGAGTAGCAG TGGTGCCATC 60

TCGGCTCACT GCAAGCTCCA CCTCCCGAGT TCACGCCATT TTCCTGCCTC AGCCTCCCGA 120

GTAGCTGGGA CTACAGGCGC CCGCCACCAT GCCCGGCTAA TTTTTTGTAT TTTTGGTAGA 180

GACGGGGTTT CACCGTGTTA GCCAGAATGG TCTCGATCTC CTGACTTCGT GATCCACCCG 240

CCTCGGCCTC CCAAAGTTCT GGGATTACAG GTGTGAGCCA CCGCACCTGG CC 292

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 262 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

TTCTTTTTTG AGACAGGGTC TTGCTCTGTC ACCCAGGCCA GAGTGCAATG GTACAGTCTC 60

AGCTCACTGC AGCCTCAACC GCCTCGGCTC AAACCATCAT CCCATTTCAG CCTCCTGAGT 120

AGCTGGGACT ACAGGCACAT GCCATTACAC CTGGCTAATT TTTTTGTATT TCTAGTAGAG 180

ACAGGGTTTG GCCATGTTGC CCGGGCTGGT CTCGAACTCC TGGACTCAAG CAATCCACCC 240

ACCTCAGCCT CCCAAAATGA GG 262

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 294 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TTTTTTTTTG AGACAAACTT TCACTTTTGT TGCCCAGGCT GGAGTGCAAT GGCGCGATCT 60

CGGCTCACTG CAACCTCCAC CTCCCGGGTT CAAGTGATTC TCCTGCCTCA GCCTCTAGCC 120

AAGTAGCTGC GATTACAGGC ATGCGCCACC ACGCCCGGCT AATTTTTGTA TTTTTAGTAG 180

AGACGGGGTT TCGCCATGTT GGTCAGGCTG GTCTCGAACT CCTGATCTCA GGTGATCCAA 240

CCACCCTGGC CTCCCAAAGT GCTGGGATTA TAGGCGTGAG CCACAGCGCC TGGC 294

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 276 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

TGACAGTCTC TCTGTCGCCC AGGCTGGAGT GCAGTGGTGT GATCTTGGGT CACTGCAACT 60

TCCGCCTCCC GGGTTCAAGG GATTCTCCTG CCTCAGCTTC CTGAGTAGCT GGGGTTACAG 120

GTGTGTGCCA CCATGCCCAG CTAATTTTTT TTTGTATTTT TAGTAGACAG GGTTTCACCA 180

TGTTGGTCAG GCTGGTCTCA AACTCCTGGC CTCAAGTGAT CCGCCTGACT CAGCCTACCA 240

AAGTGCTGAT TACAAGTGTG AGCCACCGTG CCCAGC 276

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 289 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

CGCCGGGCAC GGTGGCTCAC GCCTGTAATC CCAGCACTTT GGGAGGCCAA GGCAGGTGGA 60

TCACGAGGTC AAGAGATCAA GACCATCCTG GCCAACATGG TGAAACCCCA TCTCTACTAA 120

AAATACGAAA AAATAGCCAG GCGTGGTGGC GGGTGCCTGT AATCCCAGCT ACTCGGGAGG 180

CTGAGGCAGG AGAATGGCAT GAACCCGGGA GGCAGAAGTT GCAGTGAGCC GAGATCGTGC 240

CACTGCACTC CAGCCTGGGC AACAGAGCGA GACTCTTGTC TCAAAAAAA 289

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 298 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

AGGCTGGGCT CTGTGGCTTA CGCCTATAAT CCCACCACGT TGGGAGGCTG AGGTGGGAGA 60

ATGGTTTGAG CCCAGGAGTT CAAGACAAGG CGGGGCAACA TAGTGTGACC CCATCTCTAC 120

CAAAAAAACC CCAACAAAAC CAAAAATAGC CGGGCATGGT GGTATGCGGC CTAGTCCCAG 180

CTACTCAAGG AGGCTGAGGT GGGAAGATCG CTTGATTCCA GGAGTTTGAG ACTGCAGTGA 240

GCTATGATCC CACCACTGCC TACCATCTTT AGGATACATT TATTTATTTA TAAAAGAA 298

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 105 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

TTTTTTACAT CTTTAGTAGA GACAGGGTTT CACCATATTG GCCAGGCTGC TCTCAAACTC 60

CTGACCTTGT GATCCACCAG CCTCGGCCTC CCAAAGTGCT GGGAT 105

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 83 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

CCTCGAACTC CTAGGCTCAG GCAATCCTTT CACCTTAGCT TCTCAAAGCA CTGGGACTGT 60

AGGCATGAGC CACTGTGCCT GGC 83

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

AGAAGGTAAG T 11

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

TGGAGGTGAG A 11

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

CAGTCGTGAG G 11

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

CCGAGGTGAG C 11

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

TGGAGGTACC A 11

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

GGAAGGTCAG T 11

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

AGCAGGTGGG C 11

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

GCCAGGTACA G 11

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

TGCTGGTGAG T 11

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

ATACAGGGGAT 11

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

ATACAGGGGA T 11

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

CCCCAGGCGA C 11

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

ACGCAGTGCA A 11

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

TTTCAGATCC A 11

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

CCCCAGGAGG G 11

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

TCACAGGCTC A 11

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

CCCTAGCTCC A 11

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

CTCCAGTCCA G 11

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

TCGCAGGTGA CA 12

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

ACACAGAAGG G 11